Lógica

Estandarización de Interpretaciones

Damiano Zanardini

Graduado/a en Ingeniería Informática
Facultad de Informática
Universidad Politécnica de Madrid
damiano@fi.upm.es

Curso Académico 2010/2011

Satisfacibilidad e Interpretaciones

El problema

- Una fórmula F es insatisfacible sii no existe ninguna interpretación $\mathcal{I}$ tal que $\mathcal{I}(F) = \mathbf{v}$
- para comprobar esto, habría que analizar *todos los modelos*:
  - si F es proposicional con n proposiciones distintas, entonces hay 2^n modelos
  - en una fórmula de primer orden, las interpretaciones pueden ser infinitas!
- sería útil identificar un subconjunto de interpretaciones de F tal que
  - contenga un número *menor* de interpretaciones
  - su análisis sea suficiente para determinar la satisfacibilidad de F
- tales interpretaciones existen para cualquier fórmula, y se llaman *interpretaciones de Herbrand*

Jacques Herbrand

- (Paris, Francia, 12 de Febrero de 1908 - La Bérarde, Isère, Francia, 27 de Julio de 1931)
- doctorado en la École Normale Superieure, Paris, en el año 1929
- entró en el ejercito en Octubre de 1929
- *universo de H.*, *base de H.*, *interpret. de H.*, *estructura de H.*, *cuociente de H.*
- *Teorema de Herbrand* (dos teoremas con este mismo nombre)
- introdujo las *funciones recursivas*
- trabajó con John von Neumann y Emmy Noether
- murió escalando, por caer de una montaña de los Alpes

no exactamente como él...

Universo de Herbrand

Universo de Herbrand $H(F)$ de una fórmula F

- determina el dominio de interpretación de F en las interpretaciones Herbrand
- son todos los términos que se puedan formar usando las constantes y las funciones que aparecen en F

Universo de Herbrand: definición

$$
\begin{array}{rcl}
Const(F) & = & \text{conjunto de los símbolos de constante en } F \\
Fun(F) & = & \text{conjunto de los símbolos de función en } F \\
H_0 & = & \begin{cases} Const(F) & \text{si } Const(F) \neq \emptyset \\ \{a\} & \text{si } Const(F) = \emptyset \end{cases} \\
H_{i+1} & = & \{f(t_1, .., t_n) \mid t_j \in (H_0 \cup .. \cup H_i),\ f/n \in Fun(F)\} \\
 & = & \\
H(F) & = & H_0 \cup .. \cup H_i \cup .. \qquad \text{es el universo de Herbrand}
\end{array}
$$

Universo de Herbrand

Universo de Herbrand: ejemplos

- $F = \{p(x),\ q(y)\}$
  - $H_0 = \{a\}$
  - $H_1 = H_2 = .. = \emptyset$
  - $H(F) = \{a\}$
- $F = \{p(x,a),\ q(y) \vee \neg r(b, f(x))\}$
  - $H_0 = \{a, b\}$
  - $H_1 = \{f(a), f(b)\}$
  - $H_2 = \{f(f(a)), f(f(b))\}$
  - ...
  - $H(F) = \{a, b, f(a), f(b), f(f(a)), f(f(b)), f(f(f(a))), f(f(f(b))), ..\} = \{f^n(a), f^n(b)\}_{n \geq 0}$

Base de Herbrand

Base de Herbrand de F

- *básico*: algo sintáctico (término, átomo, cláusula) donde no aparecen variables
- *átomo básico*: átomo que se obtiene aplicando un símbolo de predicado de F a términos del universo de Herbrand de F
- la *base de Herbrand* de F es el conjunto de todos loa átomos básicos de F

Base de Herbrand: definición

$$\begin{array}{rcl} Pred(F) & = & \text{el conjunto de los símbolos de predicado en } F \\ \\ HB(F) & = & \{p(t_1, .., t_n) \mid t_j \in H(F)\,,\; p/n \in Pred(F)\} \end{array}$$

Base de Herbrand

Base de Herbrand: ejemplos

- $F = \{p(x),\ q(y)\}$
  - $H(F) = \{a\}$
  - $HB(F) = \{p(a), q(a)\}$
- $F = \{p(a),\ q(y) \vee \neg p(f(x))\}$
  - $H(F) = \{a, f(a), f(f(a)), ..\} = \{f^n(a)\}_{n \geq 0}$
  - $HB(F) = \{p(a), p(f(a)), p(f(f(a))), .., q(a), q(f(a)), q(f(f(a))), ..\} = \cup(\{\{p(t), q(t)\} |\ t \in H(f)\})$
- $F = \{p(a),\ q(y) \vee \neg r(b, f(x))\}$
  - $H(F) = \{a, b, f(a), f(b), f(f(a)), f(f(b)), ..\} = \{f^n(a), f^n(b)\}_{n \geq 0}$
  - $HB(F) = \cup(\{\{p(t), q(t), r(t, t')\} |\ t, t' \in H(F)\})$

Importante

- hablando, entre otras cosas, de Universo y Base de Herbrand, tenemos que cuidar mucho la forma de representar conjuntos

Interpretaciones de Herbrand

Una interpretación de Herbrand de F

es una interpretación $\mathcal{I}_H = (H(F), I_H)$ sobre $H(F)$ tal que:

- cada constante $a \in Const(F)$ se asocia consigo misma: $I_H(a) = a$
- cada símbolo de función $f/n \in Fun(F)$ se asocia con $I_H(f/n) = \mathcal{F} : (H(F))^n \mapsto H(F)$, tal que
  - $\mathcal{F}(u_1, .., u_n) = f(u_1, .., u_n) \in H(F)$ donde $u_i \in H(F)$
- cada símbolo de predicado $p/n \in Pred(F)$ se asocia con $I_H(p/n) = \mathcal{P} : (H(F))^n \mapsto \{\mathbf{v}, \mathbf{f}\}$, tal que
  - $I_H(p(u_1, .., u_n)) = \mathcal{P}(I_H(u_1), .., I_H(u_n)) = \mathcal{P}(u_1, .., u_n) \in \{\mathbf{v}, \mathbf{f}\}$

☞ cada átomo (básico) de $HB(F)$ tiene un valor de verdad (¿cuál? la definición no lo especifica, cada interpretación decide)

Interpretaciones de Herbrand

Interpretaciones de Herbrand: representación

Una interpretación de Herbrand se puede representar como el conjunto de los átomos básicos de $HB(F)$: afirmados si se interpretan como verdaderos, negados si se interpretan como falsos

$$\begin{array}{rcll} HB(F) & = & \{A_1, A_2, A_3, ..\} & \\ \mathcal{I}_H & = & \{A_1, \neg A_2, \neg A_3, ..\} & \text{si} \quad I_H(A_1) = \mathbf{v}, \\ & & & \phantom{\text{si} \quad} I_H(A_2) = \mathbf{f}, \\ & & & \phantom{\text{si} \quad} I_H(A_3) = \mathbf{f}, .. \end{array}$$

Terminología

- las definiciones de universo, base e interpretaciones de Herbrand se refieren a menudo a un *conjunto de cláusulas* $\mathcal{C}$, que puede ser el resultado de la estandarización de una fórmula genérica F
- en la práctica, F se considera como ya en forma clausular

Interpretaciones de Herbrand: ejemplos

- $F = \{p(x),\ q(y)\}$
  - $H(F) = \{a\}, \qquad HB(F) = \{p(a), q(a)\}$
  - hay 4 posibles interpretaciones de Herbrand:

$$\begin{array}{lcl}
\mathcal{I}_H^1 & = & \{p(a), q(a)\} \\
\mathcal{I}_H^3 & = & \{\neg p(a), q(a)\}
\end{array}
\qquad
\begin{array}{lcl}
\mathcal{I}_H^2 & = & \{p(a), \neg q(a)\} \\
\mathcal{I}_H^4 & = & \{\neg p(a), \neg q(a)\}
\end{array}$$

- $F = \{p(a),\ q(y) \vee \neg p(f(x))\}$
  - $H(F) = \{f^n(a)\}_{n \geq 0}, \qquad HB(F) = \cup(\{\{p(t), q(t)\} |\ t \in H(F)\})$
  - hay un número infinito de interpretaciones de Herbrand

$$\begin{array}{lcl}
\mathcal{I}_H^1 & = & \cup(\{\{p(t), q(t)\} \mid t \in H(F)\}) \\
\mathcal{I}_H^2 & = & \{p(a)\} \cup \{\neg p(t) \mid t \in H(F) \setminus \{a\}\} \cup \{q(t) \mid t \in H(F)\} \\
\mathcal{I}_H^3 & = & \{p(t) \mid t \in H(F)\} \cup \{\neg q(t) \mid t \in H(F)\} \quad \ldots
\end{array}$$

Interpretaciones de Herbrand

Instancias básicas

Una *instancia básica* de una cláusula es una cláusula, que se obtiene sustituyendo las variables de la cláusula original con términos de su universo de Herbrand

- es una aplicación universal
- a través de una interpretación de Herbrand se puede dar un valor de verdad a una fórmula a partir del valor de verdad de sus instancias básicas

Ejemplo: $F = \{p(a),\ q(b) \vee \neg p(x)\}$

- $H(F) = \{a, b\}$ $\qquad HB(F) = \{p(a), p(b), q(a), q(b)\}$
- $\mathcal{I}_H = \{p(a), \neg p(b), q(a), \neg q(b)\}$
- primera cláusula: verdadera porque su única instancia $p(a)$ es verdadera en $\mathcal{I}_H$
- segunda cláusula: falsa porque una instancia, $q(b) \vee \neg p(b)$, es verdadera en $\mathcal{I}_H$, pero la otra, $q(b) \vee \neg p(a)$, es falsa

siendo F la conjunción de ambas cláusulas, es falsa para $\mathcal{I}_H$ (ya veremos por qué)

Interpretaciones de Herbrand

$\mathcal{I}_H$ correspondiente a $\mathcal{I}$

Dada $\mathcal{I} = (D, I)$, una interpretación de Herbrand $\mathcal{I}_H = (D_H, I_H)$ *corresponde a* $\mathcal{I}$ para F si satisface a la siguiente condición:

- I' es una mapa total de $H(F)$ a D tal que
  - $I'(c) = d$ si $I(c) = d$ (constantes)
  - $I'(f(t_1, .., t_n)) = \mathcal{F}(I'(t_1), .., I'(t_n))$ donde $I(f/n) = \mathcal{F}/n$
- para cada átomo básico $p(t_1, .., t_n) \in HB(F)$, $I_H(p(t_1, .., t_n)) = \mathbf{v}$ (resp., $\mathbf{f}$) si $I(p)(I'(t_1), .., I'(t_n)) = \mathbf{v}$ (resp., $\mathbf{f}$)

☞ esta definición puede parecer inútilmente complicada, pero...

Interpretaciones de Herbrand

Ejemplo: $F = \{p(x), q(y, f(y, a))\}$, $D = \{1, 2\}$

- $I(a) = 2$
- $I(f/2) = \mathcal{F}/2$: $\quad \mathcal{F}(1,1) = 1 \quad \mathcal{F}(1,2) = 1 \quad \mathcal{F}(2,1) = 2 \quad \mathcal{F}(2,2) = 1$
- $I(p/1) = \mathcal{P}/1$: $\quad \mathcal{P}(1) = \mathbf{v} \quad \mathcal{P}(2) = \mathbf{f}$
- $I(q/2) = \mathcal{Q}/2$: $\quad \mathcal{Q}(1,1) = \mathbf{f} \quad \mathcal{Q}(1,2) = \mathbf{v} \quad \mathcal{Q}(2,1) = \mathbf{f} \quad \mathcal{Q}(2,2) = \mathbf{v}$

En este caso I' es igual que I (sobre $H(F)$)

- $I_H(p(a)) = I(p(a)) = \mathcal{P}(I(a)) = \mathcal{P}(2) = \mathbf{f}$
- $I_H(q(a,a)) = I(q(a,a)) = \mathcal{Q}(I(a), I(a)) = \mathcal{Q}(2,2) = \mathbf{v}$
- $I_H(p(f(a,a))) = I(p(f(a,a))) = \mathcal{P}(\mathcal{F}(2,2)) = \mathcal{P}(1) = \mathbf{v}$
- ...

Interpretaciones de Herbrand

Varias interpretaciones de Herbrand

Puede haber varias $\mathcal{I}_H$ correspondientes cuando F no tiene constantes. En este caso, I no interpreta H_0 (es decir, $I' \neq I$) y por tanto la interpretación de $a \in H_0$ es arbitraria para I_H:

- $F = \{p(x)\}, \;\; D = \{1, 2\}, \;\; p(x)$ significa que x es par
- $H(F) = \{a\}, \;\; HB(F) = \{p(a)\}$
- $I'(a) = 1$ y $I'(a) = 2$ están las dos permitidas
- $\mathcal{I}_H^1 = \{\neg p(a)\}$ suponiendo $a \rightsquigarrow 1$
- $\mathcal{I}_H^2 = \{p(a)\}$ suponiendo $a \rightsquigarrow 2$

Lema

Si la interpretación $\mathcal{I} = (D, I)$ satisface F, entonces todas las interpretaciones de Herbrand de F correspondientes a $\mathcal{I}$ también satisfacen F

- *ej:* $F = \forall x p(x) \wedge \forall x q(f(x))$

Interpretaciones de Herbrand

Teorema

Una fórmula F es insatisfacible sii es falsa para todas sus interpretaciones de Herbrand

Demostración ($\longrightarrow$).

1. F es insatisfacible
2. es falsa para todas las interpretaciones sobre todos los dominios
3. en particular, todas las interpretaciones de Herbrand la hacen falsa

Interpretaciones de Herbrand

Teorema

Una fórmula F es insatisfacible sii es falsa para todas sus interpretaciones de Herbrand

Demostración ($\longleftarrow$).

❶ F es falsa para todas las interpretaciones de Herbrand

❷ supongamos que F sea satisfacible

❸ hay una interpretación $\mathcal{I}$ que satisface F (por ❷)

❹ por el lema anterior, las interpretaciones de Herbrand correspondientes también satisfacen a F

❺ contradicción entre ❶ y ❹, entonces ❷ es falsa

❻ F es insatisfacible (por ❺)

Interpretaciones de Herbrand

En la práctica

Para estudiar la satisfacibilidad de una fórmula F, basta con estudiar las interpretaciones de Herbrand de su forma clausular $FC(F)$

Para cada interpretación de Herbrand de $FC(F)$

- calcular las instancias básicas de las cláusulas
- asignar un valor de verdad a cada instancia

☞ $FC(F)$ es verdadera en $\mathcal{I}_H$ sii *todas* las instancias básicas de *todas* las cláusulas son verdaderas en $\mathcal{I}_H$

☞ F es satisfacible sii *alguna* interpretación de Herbrand evalúa $FC(F)$ como verdadera

Ejemplo: $F = \{p(x),\ q(y)\}$

- $H(F) = \{a\}$ $HB(F) = \{p(a), q(a)\}$

Hay 4 interpretaciones de Herbrand

- $\mathcal{I}_H^1 = \{p(a), q(a)\}$
- $\mathcal{I}_H^2 = \{p(a), \neg q(a)\}$
- $\mathcal{I}_H^3 = \{\neg p(a), q(a)\}$
- $\mathcal{I}_H^4 = \{\neg p(a), \neg q(a)\}$

Instancias básicas: $\{p(a), q(a)\}$

- $\mathcal{I}_H^1$ es un modelo porque hace verdad las dos instancias
- $\mathcal{I}_H^2$, $\mathcal{I}_H^3$ y $\mathcal{I}_H^4$ son contramodelos porque falsifican al menos una instancia

Por tanto, F es satisfacible

Interpretaciones de Herbrand

Ejemplo: $F = \{p(y),\ q(a) \vee \neg p(f(x)),\ \neg q(x)\}$

- $H(F) = \{f^n(a) \mid n \geq 0\}$
  $HB(F) = \{p(t) \mid t \in H(F)\} \cup \{q(t) \mid t \in H(F)\}$

Hay infinitas interpretaciones de Herbrand. Por ejemplo

- $\mathcal{I}_H^1 = \{p(t) \mid t \in H(F)\} \cup \{q(t) \mid t \in H(F)\}$
- $\mathcal{I}_H^2 = \{q(a)\} \cup \{\neg q(t) \mid t \in H(F) \setminus \{a\}\} \cup \{p(t) \mid t \in H(F)\}$

Instancias básicas

$$\begin{array}{rcl} p(y) & \rightsquigarrow & p(a), p(f(a)), p(f(f(a))), .. \\ q(a) \vee \neg p(f(x)) & \rightsquigarrow & q(a) \vee \neg p(f(a)), q(a) \vee \neg p(f(f(a))), .. \\ \neg q(x) & \rightsquigarrow & \neg q(a), \neg q(f(a)), .. \end{array}$$

Todas las interpretaciones de Herbrand falsifican al menos una instancia, así que F es insatisfacible

El Teorema de Herbrand

Teorema (Teorema de Herbrand)

Un conjunto de cláusulas $\mathcal{C}$ es insatisfacible sii existe un conjunto finito e insatisfacible de instancias básicas de cláusulas de $\mathcal{C}$

- la demostración (que normalmente se divide en dos partes) usa la técnica de los *árboles semánticos*, que este año no hacemos
- pero el enunciado, hay que saberlo

El teorema sugiere un método

Dado un conjunto $\mathcal{C}$ de cláusulas, generar instancias básicas incrementalmente, y meterlas en un conjunto hasta que dicho conjunto se vuelva insatisfacible:

$$
\begin{array}{l}
B = \emptyset; \\
\text{mientras } (B \text{ es satisfacible}) \\
\quad b = \text{nueva-instancia}(\mathcal{C}); \\
\quad B = B \cup \{b\};
\end{array}
$$

Implementaciones del teorema de Herbrand

Se necesita una *estrategia* para (1) generar instancias y (2) decidir su satisfacibilidad

- *método de Gilmore (1960)*
- *método de Davis-Putnam (1960)*
- método de Resolución de Robinson (1965)